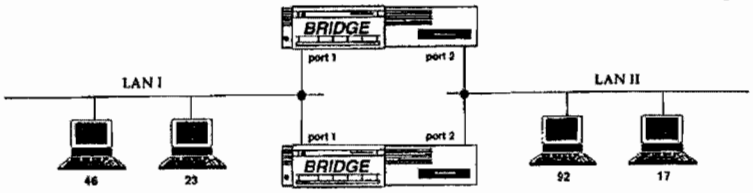
2007 IIB 4B15 Dr T D Wilkinson
tdw@eng.cam.ac.uk

4B15 Crib 2007 (answers here are longer than expected)

Q3 a) LANs tend to 'evolve' from previous incarnations. Because of this process of evolution, the topology or interconnection of the LANs and stations could be sub-optimal, which often leads to loops in the network. So how can this happen?

- Historically bridges were slow, so the temptation was to double up bridges to speed up network operation.
- Complex networks are often very difficult to keep tabs on especially over several years of change and evolution.
- Redundancy is often added to a network by planners to guarantee bandwidth at a given cost.



If we consider the above bridge connection, using ordinary transparent store and forward bridges which have up to date address tables, what will happen when frames are sent? If station 46 sends a frame to station 17, each bridge receives the frame and sends it to the appropriate port (port 2) and after queuing, the frames appear at station 17. Station 17 will receive two frames from station 46 which violate the non-duplication invariant. In fact every station on LAN II will receive duplicate unicast frames from LAN I. If LAN 46 sends a multicast frame, both bridge A and B will forward the frame, however bridge A will receive the forwarded frame from bridge B (and vice versa) on port 2. Upon receiving this frame, the bridge will look at the source address of the frame and reconfigure station 46, thinking it is now connected to port 2. This process will continue indefinitely with both bridges continually updating their address tables.

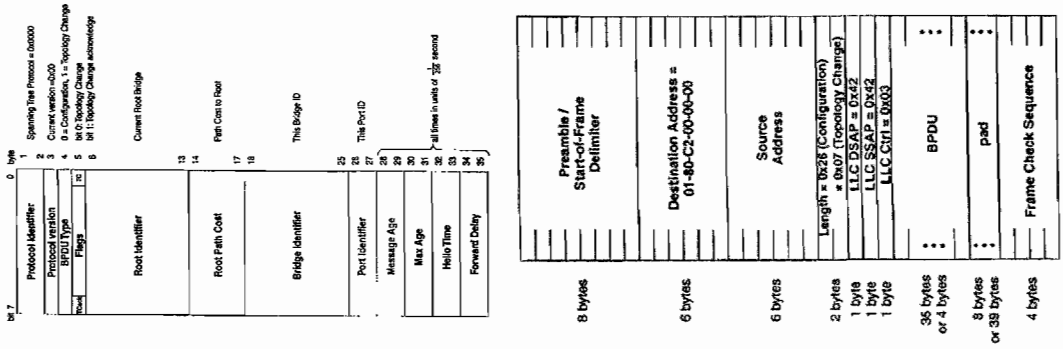
b) The STP operates on the principle that all designated bridges (including the root) advertise their current understanding of the spanning tree and their internal state by emitting, on a regular basis, through their designated ports, configuration messages (encoded as bridge protocol data units (BPDU)). All bridges listen to these configuration messages and compare them with their own internal information. When a bridge internally feels it has a better claim to be root or a designated bridge, it initiates a topology change. The regular transmission of these control messages maintains the steady state. If a link fails and the messages stop, then previously inactive ports may become active to re-route data. In the steady state, the STP follows:

1. Once every Hello time (usually 2 seconds), the root bridge transmits a configuration message encoded as a BPDU. This identifies the root bridge.
2. All bridges sharing links with the root bridge receive the message and process it internally. BPDUs are never forwarded.
3. Designated bridges (or those preparing to be designated) use the information received from the root bridge, and update the identifier, path cost and port identifier and then transmit it out of its designated ports.
4. This process will repeat from bridge to bridge until there are no bridges left before the stations.

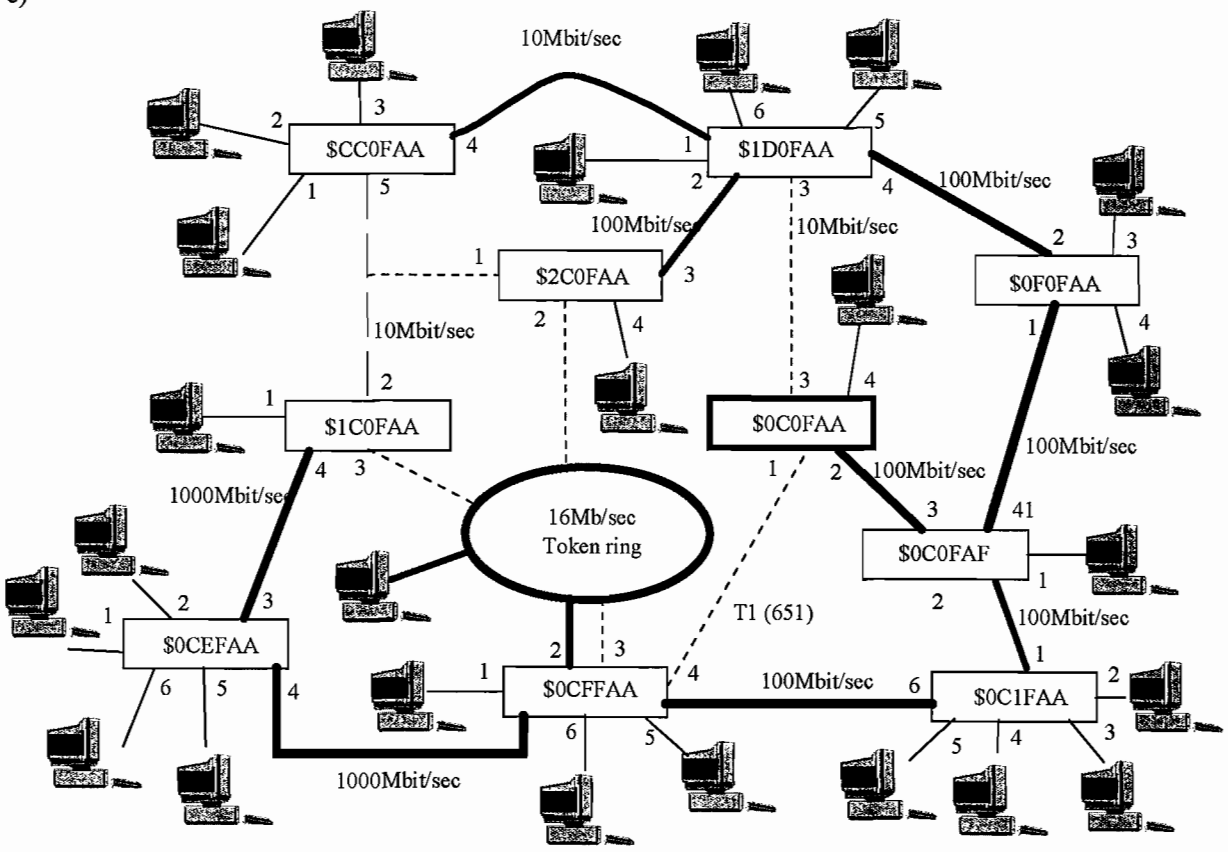
Every bridge receiving the configuration messages compares the information received with its own internal state and knowledge. Specifically the bridge compares:

- The root identifier with its own identifier. If it is numerically lower then it initiates topology change and transmits configuration messages with itself as root.
- The path cost in received messages to the path cost available to this bridge through any other ports. Hence if it can offer a lower cost path, then it initiates a topology change. If costs are equal then it compares bridge identifiers or port identifiers as required.

When transmitted on a LAN, BPDUs are further encapsulated in MAC frames using LLC type 1, with a DSAP and SSAP of 0x42. The MAC source address is the MAC address of the port through which the frame is being transmitted. The MAC destination address is the multicast address 01-80-C2-00-00-00. The use of a multicast address means that the source bridge does not need to know the destination address of other bridges.

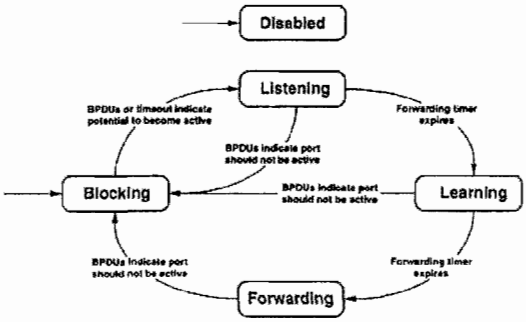


c)



d) Each bridge port can be in one of five possible states.

- **Disabled.** When disabled, a port will neither receive nor transmit frames or STP messages. This is usually due to a port fault or duff link.
- **Blocking**. A port that is enabled but is neither designated nor root port will be in the blocking state, which is in standby or not currently required by the STP. It will not receive or transmit data frames nor transmit BPDUs, but it will listen for others BPDUs (or time out if none heard) to determine if and when the port should be consider becoming active in the spanning tree. The blocking state is usually entered on power up. Upon learning (by either BPDUs or time-out) that it needs to become active, it will go through three stages: listening, learning, forwarding.
- Listening. In this state, the port will not forward data, but is listening to (and possibly sending) BPDUs to determine the spanning tree. If the port decides it is not suited to becoming designated it will return to the blocking state.
- Learning. In this state, the port is preparing to start forwarding data. Since the ports address table will be empty, it needs to wait (known as the forwarding delay) as a port with an empty table will flood all data frames.
- Forwarding. Once the bridge has spent time learning addresses, it is allowed to forward data frames. This is the steady state for an active port on the spanning tree.



e) The STP has become a very important part of modern WANs and MANs as it offers a series of fixed connection paths through a combination of networks. Even though layer 3 routers can be very sophisticated, the processing overhead of this sophistication is very high which leads to excessive latency across multiple nodes or networks. If a ST can be set up, then there is no decision required as each path is a single route across the catenet ad can be very fast and low latency. This ties in with older ideas such as source routing and frame relay across mesh networks where the meshes are defined by the branches of the ST.

A second advantage of the ST is that is can adapt when changes occur as it is continuously updating itself and monitoring its own structure. This means that an entire internetwork can self optimise rather than just a series of links.

The problem is compatibility between manufacturers to get them all to follow the same STP rules.

Q1 a) *Connection Oriented.* In this case, the link provides both error and flow control between communicating stations. This sort of service requires a complex interaction between stations, with data stored at both ends on the structure and sequence of frames. There is usually a procedure for call setup and take down established as well. This is more robust and reliable, but there is no such thing as a free lunch.

An example of a CO network is X25 which uses HDLC calls to maintain a connection or virtual channel. A CO network is ideal for voice data as it can offer a fixed connection with a suitable service to minimise delay. It is not efficient for data as it will hold the VC even if data is not transmitted and will have to set up the VC each time.

Connectionless. This is the case when service is on a best efforts basis, with no guarantees of transmission or data recovery mechanisms. Errors can be detected and frames prevented from transmission to higher layers, but there is no system for retransmission. There is no feedback between stations as to the success or failure of frame transmission. This is simple, but data will be lost on even the most reliable channels.

A connectionless network is IP and the internet. Packets are sent via different routes on a best efforts basis. This is good for data, as it suits the bursty nature, but is not good for guaranteeing delay with voice.

b) The basis of the Internet is store-and-forward packet switching where the packet is buffered until fully received and error checked before being processed either by the fast path for data packets or on the edge of the fast path for more complex packets. Based on this process, the network layer should provide the services with the following goals in mind, to the transport layer.

1. The services should be independent of router technology.
2. The transport layer should be shielded from the number, type and topology of the routers.
3. The network addresses should use a uniform numbering plan, even across LANs and WANs.

This offers a considerable amount of freedom, however there is one battle which has raged over the history of networks: whether the network layer should offer connection oriented or connectionless service. The Internet computer community argues (based on 20 years of network function) that the subnet is inherently unreliable and that the routers should control the flow of packets, control errors etc. This means a connectionless service is required with an in built address structure which must be carried by every packet. This, however means that quality of service (QoS) is inherently a challenge.

On the other hand, the telecommunications networks providers believe that (based on 100 years of telephone service) the way forward is a connection oriented service where a virtual circuit is established before data is transmitted. They believe that QoS is the key and very difficult to provide without connections in the subnet, especially for voice and video. With voice over IP (VoIP), TV channels and video on demand looking like the next big services, even the Internet is having to acknowledge the use of connection oriented services.

c) ICMP is an attempt to allow simple control features to be implemented within a native IP format. It allows the use of features such as acknowledgements and various sequencing and control mechanisms. It is normally implemented in the IP packet header and is indicated by the presence of the option field in the packet. When it is included, it extends the length of the header due to the extra information in these fields. If is similar to LLC in layer 3 systems.

There are 2 main issues with using ICMP. The first is compatibility, as all routers which use it must understand and obey the commands used in the header. This is possible when the network equipment is all made by the same manufacturer and is all of the same generation, however this is rarely the case. Older routers will not be ICMP compatible and will no be able to implement its features. Also there are many different variant of ICMP so it is difficult to guarantee that all routers will be using the same commands in the same way. The second issue is that the use of the options field will the packet through a router that uses the fast path as this the options field is used as a path parsing filter. Hence ICMP field packets will be delayed if other fast path packets arrive which contradicts the purpose of the ICMP in the first place.

d) It is possible to configure and operate a network, which has no knowledge of its interconnection or the positions of stations upon it. All of the route configuration and set-up processing is performed by the stations at either end of the route or link(s). This is referred to as *source routing*. The logic behind it was that source routing required more computing power at the stations. There is no source routing mechanism for Ethernets (Xerox and IBM didn't get on). There are a whole host of different arguments about link efficiency, traffic management, link redundancy, latency control etc which can be used to justify source routing, however an important lesson is it parallels with the switching protocols used in the synchronous digital hierarchy (SDH) and asynchronous transfer mode (ATM) switching.

Each end station in the catenet connection must learn all about the available paths in the catenet through the process of route discovery. There may be multiple paths through a catenet and the stations must select the route to use. Normally both stations will use the same route forward and in reverse, with the route being stored in cache until the connection is over. During route discovery, the station may learn that the end station is on the same ring in which case normal token procedure is invoked and source routing not used. During the discovery process the end stations will also learn the maximum transmission unit (MTU) which can be handled across the path, which will then set the MTU used in data transmission.

Source routing inserts extra fields into the MAC frame, hence it is necessary to indicate presence of a source routed frame. This is done by using the first bit of the 48bit source address in the MAC frame, which normally defines unicast or multicast address format. This bit is normally wasted as multicast source address is meaningless, hence it is used to indicate (set to 1) a source routed frame. This is a good filtering bit for LANs such as Ethernet where source routing is not permitted.

A key issue in source routing is the length of time that routes are stored in the stations, as every route discovery takes time. This is not very efficient.

e) MPLS works by grouping together packets with the same destination. These labelled packets then all follow a common route to increase the performance across a set network or tunnelling through another network. Source routing could be used to find these suitable routes for the packet groups before sending through the network. This is a route discovery process which is then optimised and translated into a suitable label for the group. The process of source routing would be implemented at the ingress point in the MPLS network.

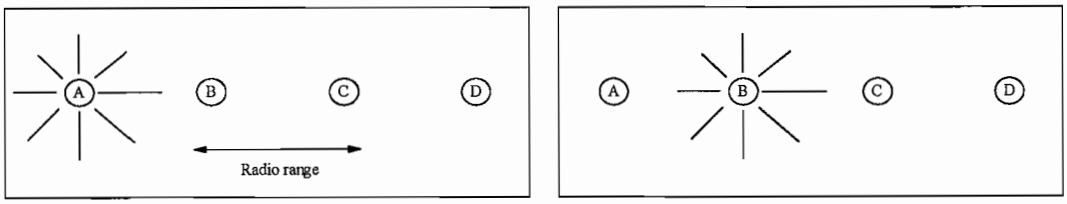
Q2 a) The carrier sense multiple access with collision detection (CSMA/CD) protocol is a contention protocol. On a CSMA/CD LAN the terminals do not request permission from a central controller before transmitting data onto the transmission channel; they contend for its use. When a frame is queued, ready for transmission, the sending station 'listens' to check whether the physical channel is in use by another station, referred to as *sensing carrier*. If it is in use, then the station *defers* and it waits before trying transmit its frame. Following the transmission in progress, (ie when the carrier is no longer sensed), the station waits a period of time known as the interframe gap to allow the physical channel to stabilise and to additionally allow time for receivers to reset. The station then begins to transmit. If this is the only station with a frame queued, then it is transmitted without interruption. Once this is done, it will begin to transmit the next frame in the queue.

Even when it starts to send a frame, the station needs to continue checking the path to make sure that no other stations have started sending data at the same time. If the sending terminal's output does not match that which it is simultaneously monitoring on the transmission path, it knows there has been a collision. In the event of a collision, all involved stations continue to transmit for a short time so that all parties are aware, in a process known as *jamming*. They then abort the remainder of their frames and wait for a random period of time known as *backing off*, before beginning the transmission of the frame again. If a frame encounters 16 collisions, then it is discarded and the next frame attempted. The back off time for any retransmission attempt is a random variable with an exponentially increasing range for repeated transmission attempts. The random variable r selected on the nth transmission attempt of a given frame is $0 < r < 2^k$ where k = MIN (n,10).

The back off time is measured in units of the worst-case round trip propagation delay of the channel or *slotTime*. This is 512 bit times for all rates except 1Gbit/s where it is 4096 bit times. Eg, 51.2μsec (10Mbit/sec), 5.12μsec (100Mbit/sec) and 4.096μsec (1Gbit/s). To receive data, the medium access control (MAC) software in each station monitors the transmission path, decoding the destination address of each frame passing through to find out whether it is the intended destination. If it is, the data is error checked with the frame check sequence and then passed ton the higher layers, otherwise the frame is discarded. The receiver rejects any frame shorter than the slotTime.

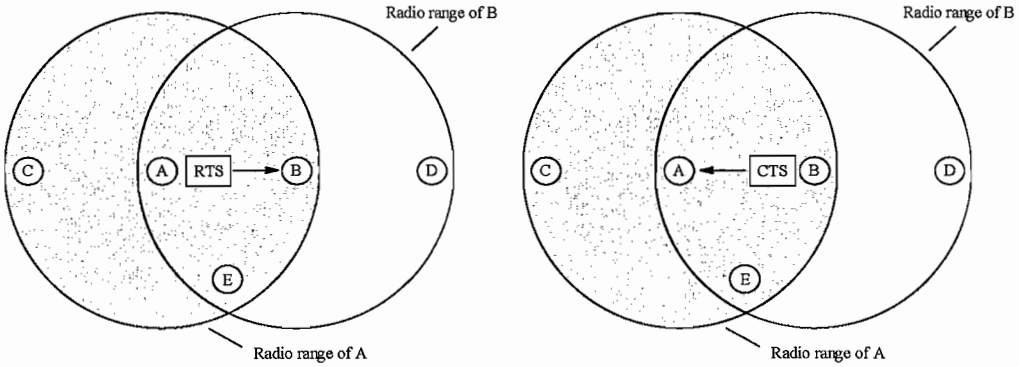
Theory suggests that random collision of a large number of devices can lead to transmission degradation under heavy traffic, however traffic is rarely random as most LAN transmissions involve a central server system. Traffic problems can be alleviated by aggregation or subdivision into smaller LANs.

b) Initially, early WLANs were run using CSMA/CD quite efficiently until it was realised that there were occasional miss-transmissions and blocked links dropping frames. This problem was due to the use of CSMA/CD in conjunction with the wireless physical layer.



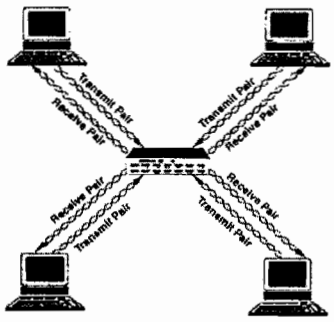
If we consider the above scenario, where four stations are in the local area (they could be base stations or mobile stations). The radio range is such that A and B are within each other range. C can possibly communicate with B and D but not A. When A is transmitting to B and C senses for a carrier, it will not hear A and will start transmitting as well, corrupting the original frame from A. This is called the hidden station problem. The solution is to use multiple access with collision avoidance (MACA), which became the basis for the IEEE standard 802.11. In MACA, the sender stimulates the receiver to send a short frame which alerts other stations nearby of its impending activity and prevents collisions.

4B15 Crib 2007 (answers here are longer than expected)



Consider the stations above. A sends a frame to B, known as a request to send (RTS) frame, which is a 30 byte frame which contains the length of the data to follow. B then replies with a clear to send (CTS) frame which also contains the data length. When A gets the CTS it starts transmitting data. Any station hearing the RTS from A must be quiet until the CTS from B has been sent. Any station close to B will hear the CTS and must remain silent for the length of the data frame. C is close to A, but not B so it must wait until the CTS has reached A, then it is free to transmit. D is close to B, but not A so it must wait until the data has been sent to B. E is in range of both A and B so it must be silent for the whole operation. Collisions can still occur if RTS frames are sent at the same time, hence colliding stations will use the same back off procedure as CSMA/CD. Later improvements were added such as acknowledgements and CSMA to improve the protocol (known as MACAW)

- c) A means of operation of a modern Ethernet is *full duplex*, where each direction of traffic to or from a station occupies its own physical channel. This has the effect of negating MAC protocols such as CSMA/CD and token rings. There are two main factors which allow full duplex operation.

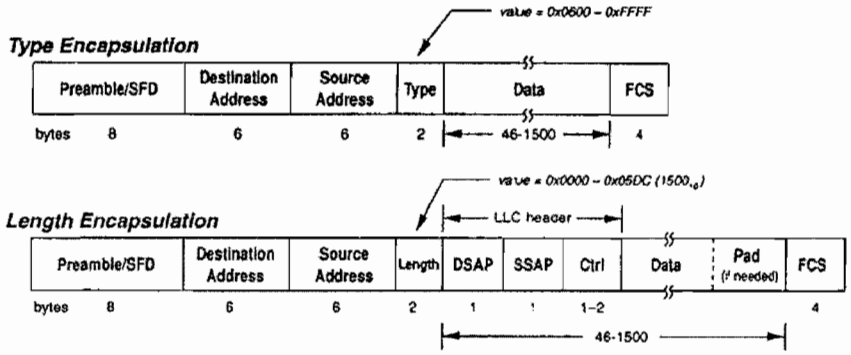


- The use of dedicated media such as structured cabling
- The use of microsegmentation about a switch.

Twisted pair based Ethernet systems such as 10baseT, 100baseTX and 1000baseT etc can, in principle, support full duplex operation, whereas token ring and FDDI have always supported full duplex. The use of dedicated media allows the deployment of switching hubs (ie bridges) rather than repeaters at the centre of star wired systems. Whereas with a repeater, all stations connecting to the hub share the available channel and have to arbitrate for access, with a switching hub each of the attached stations have a dedicated link between itself and the hub. In order to implement a full duplex link efficiently, certain aspects of the MAC protocol must be deactivated as they are no longer useful. For the case of Ethernet:

- Deactivate the carrier sense function as no-one else will be using the link.
- Deactivate the collision detect function as there cannot be collisions.
- Disable the loop back of transmitted data as the first two are gone.

d) An Ethernet frame can take two possible forms. The preamble/SFD, address and frame check sequence (FCS) are common to both types. They are referred to as *type* and *length encapsulation* frame formats.



As Ethernet has evolved it has become increasingly popular and is used almost everywhere in the world and the loss of CSMA/CD has made it basically a standard frame format with few other used features (such as LLC now left to higher layers). The common usage has meant that most hardware must convert to and from Ethernet frames when using a WAN or MAN, hence it is simplest just to use the frame itself and maybe a simple header such as in PPP. The main drawback is that the MTU due to CSMA/CD is set at 1500 bytes (due to the old usage of coaxial cable) has evolved with this process. This is not an ideal limit on packets at the MAN and MAN level. It leads to complexity especially if MPLS is used as well.

e) Manufacturers also actively market layer 4 switches which operate on the transport layer, normally providing end-station to end-station services across an underlying internetwork. An example of this could use the transport control protocol (TCP) to provide reliable error and flow controlled connection oriented services across a connectionless oriented network. TCP operates only between end stations and underlying bridges and routers are not involved. Strictly speaking a layer 4 switch is not possible as there is no means of identification at this level. However higher layer policies such as specific address or domain filtering and security could be implemented. Many higher level features can be specified for the switch to operate on such as network management, congestion control and delay sensitive traffic priority such as video streaming. The control of layer 4 operation using protocols such as TCP is often done on a stream of packets referred to as an *application flow*. Flow control is a very sophisticated feature of TCP systems and is a current hot topic of academic research.

The majority of TCP control is done through short packets of 40 bytes which implement the control and flow functions. This is at the shortest length limit of Ethernet, and often care is needed by padding to avoid packets being discarded or ignored by the Ethernet hardware. TCP could work better with longer packets when sending data as it assigns flows across a window which combines multiple packets. For short data fragments his is Ok, but for longer data transfers, the process becomes less efficient and can disrupt the performance. A longer or variable frame length would be much better.